Data Visualization and Evaluation for Industry 4.0 using an interactive k-Means Algorithm

Dieter Meiller

East Bavarian Technical
University Amberg-Weiden
Kaiser-Wilhelm-Ring 23
92224, Amberg, Germany

d.meiller@oth-aw.de

Florian Niewiera

East Bavarian Technical
University Amberg-Weiden
Kaiser-Wilhelm-Ring 23
92224, Amberg, Germany

f.niewiera@oth-aw.de

ABSTRACT

The project ISAC@OTH-AW will focus on developing an innovative expert system for data visualization and optimization to produce better manufacturing processes. A mandatory part of the project is the appropriation of industry 4.0 technology benefits, like efficiency, quality and time to market rate, for **SME**s (**S**mall and **M**edium **E**nterprises).

Keywords
Industry 4.0, swarm intelligence, swarm visualization, interactive k-Means, particle swarm optimization, big data, information visualization, visual analytics, machine learning.

1. INTRODUCTION

Modern digital factories are monitored and remote controlled. Large amounts of data are transmitted to interfaces for controlling and monitoring. The next stage is to analyze the data in order to make predictions. The modelling of algorithms for analysis can be supported by interactive visualizations. This topic will be focused in this project.
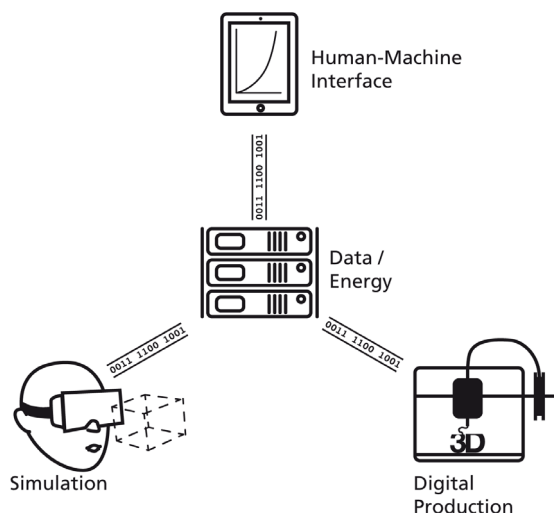

Figure 1. Structure of ISAC@OTH-AW

Four main topics will be addressed within the ISAC@OTH-AW project (Industry Software Application Center at East Bavarian Technical University Amberg-Weiden) :

1. The improvement of production cycles and process workflows, which are based on characteristic production data.

2. Opportunity of facility optimization by virtual plant simulation.

3. Intelligent networked and power supplied controllers and machines.

4. Innovative visualization design for process monitoring and the optimization of data analysis for big data sources.

Four project cooperation partners of two faculties at the Technical University will handle the mentioned topics. The first results are presented in this short paper. They aim to portray the different aspects of human-machine-interface, visualization and the analysis of process data for production optimization in subproject number four.

2. Current requirements in industrial environments by small and medium enterprises (SME)

The desire and need by companies for an optimization, visualization and handling of big data appearances is greater than ever before. Industry 4.0 technologies can eliminate many frequently occurring problems and ease process procedures in a SME. This will be adapted and refined in the

ISAC@OTH-AW project. These problems include many helpdesk center situations for common customer support. The innovative storage and analysis of costumer queries could offer fast and more reliable solutions that are unique to each problem for an individual support. Hence, the reduction of customer support time would raise efficiency and could lead to a higher customer satisfaction with less effort for employees.

Another indispensable operation field for industry 4.0 benefits is the enhancement and optimization of production sequences and machine parameters for example CNC milling, drilling, spray casting, 3d printing. The improvement of these parameters would help accomplish many things including faster process times or a conservative tool usage.

Another possible application is the handling and long term archiving of big data occurrences for selected analysis of relevant data parts in combination with innovative and informative visualization methods. The topic in this periphery is the identification of undetected process correlations, the gain of a clearly represented production overview or an easier detection of essential information.

Furthermore, even data evaluation for logistical problems could benefit from these optimizations to reducing operation time and avoiding an excess use of space and resources.

3. State of the art

As mentioned, the main targets of the analysis component in this project are self-clustering and evaluation algorithms. In particular, big data pools with a variety of different data types will be processed. In the first step, intelligent models and algorithms analog to swarm visualization [Mei15a] and particle Swarm Optimization [Mal15a] [Sha12a] [Kuo11a] will complete the self-organization of these data objects. After this extensive phase of organization, all data objects will be analyzed by using machine learning algorithms [Na15a]. The obtained results of the unsupervised machine learning supported analysis shall deliver new insights.

An advantage of visualizing an amount of m objects in a swarm is, that the provided data could have n features where n can be greater than three. Hence, the dimensionality n is higher than it is suitable for mapping the features to an axis for the drawing. A common way to visualize correlations between data examples is to reduce the dimensionality of the multivariate data using the Principal Component Analysis (PCA) algorithm [Joll02a]. This approach is not suitable when the amount m is big or varying, because the algorithm is expensive to perform. Hence, the data will be visualized as a swarm, where single objects attract each other when they have

similar features [Mei15a]. In this approach it is not necessary to reduce the dimensionality because the similarity can be computed from more than three features. The strength of the similarity will increase or decrease the attraction between two single objects and can be observed in a 2D space. This can come to use when performing tasks with industry data. An example of this is to detect classes of support requests in the help desk and to route them to the right administrator. The problem with applying machine-learning techniques such as the k-Means algorithm to identify clusters [Hart79a] is the choosing of the number of clusters K, where K should be smaller than m. This could be solved by visualizing the data and estimate the value of K. Thus, swarm visualization could be an approach to support a subsequent clustering algorithm like k-Means.

4. Procedure model

The computation of the clustering algorithm (e.g. k-Means [Eri11a] [Esm13a]) could be mapped into the swarm visualization as well: Each of the K cluster centroids μ_k ($k \le K$) could be visualized as an object in the swarm, because it has the same dimensionality as the data objects and will attract similar data. Additionally, clustering algorithms (e.g. k-Means) could be run interactively. An operator could be able to add and remove cluster centroids at runtime, simply by touch:

```
1.   K := 0;
2.   Repeat while clustering
3.      if touch on x⁽ⁱ⁾
4.        K := K + 1;
5.        add Centroid μK := x⁽ⁱ⁾
6.      endIf
7.      if touch on μk
8.        K := K – 1
9.        remove Centroid μk
10.   endIf
11.   //Assign Data to centroid
12.   for i := 1 to m
13.      C⁽ⁱ⁾ := index k of
                μk most similar to x⁽ⁱ⁾
14.   endFor
15.   //Make centroid similar
        to assigned data
16.   for k := 1 to K
17.      n := 0
18.      for i := 1 to m
19.        if C⁽ⁱ⁾ = k
20.          μk := μk + x⁽ⁱ⁾
21.          n := n + 1
22.        endif
23.        μk := μk / n if n > 0
24.      endFor
25.   endFor
26. EndRepeat
```

Algorithm 1. Interactive k-Means algorithm

In Algorithm 1 a new interactive version of the k-Means algorithm is described. $C^{(i)}$ is an index ($i \leq m$) and $x^{(i)}$ is one data sample, a vector of n features ($x^{(i)} \in \mathbb{R}^n$). The algorithm runs continuously.

An endless loop encloses the procedure (line 2-26) to make it run as long as the operator needs to find clusters. At the start, there are no cluster centroids present (number of centroids K is 0 at line 1). The user is able to touch (click) on a representation of a data object $x^{(i)}$ to add a new cluster centroid at this location. The centroids will be removed (lines 8-10) by repeatedly clicking on them. After adding one more centroid, the cluster algorithm will (re-) assign the data points to all centroids available (lines 12–14). In lines 16–25 the centroids will be matched to the data points they correspond with. After that, the centroids will move ahead into swarm visualization to the data points which are assigned to them.

Since we are still in the initial phase of the ISAC@OTH-AW project the algorithms have not been fully developed or chosen yet. This is only an example of how (swarm-) visualizations can support methods of machine learning. We plan to support further techniques of machine learning with swarm visualizations: methods for anomaly detection to determine defects in manufacturing, supervised learning with support vector machines and neuronal networks to predict good machine configurations could benefit from support by interactive visualizations as well.

5. Approach and current status of the implementation

Actual production data from different facility machines like CNC milling machines is collected and accumulated in the first step of the project. This information comprised e.g. timestamps, tool position data, speed-up parameters. All information is saved into a real time compatible NoSQL database like MongoDB. It is an open source, cross platform document oriented database. It is possible to collect different types of data in great quantities because of schema free file storage. Additionally, static information (tool parameter, specimen parameter) belonging to these dynamic machine information parameters will be stored to the data base.

In a second step the first realization and implementation of a purpose build visualization and data analysis framework was initiated. For this, the framework "DataOcean@ISAC" (Figure 2) was created. The application visualizes all process data and parameter as swimming particles in a barrier free environment, such as a fish swarm in the ocean. These individual data objects are self-organized and behave like a fish in a swarm. Algorithms based on Flock Models [Cui00a] [Fow00a], Swarm

Intelligence [Cui00c] combined with k-Means [Bin00a] and Particle Swarm Optimization [Cui00b] will be used to organize the data structures by themselves.
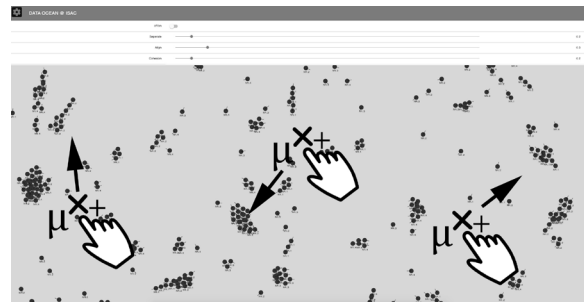

Figure 2. Process of adding centroids to intelligent swarm visualization (DataOcean@ISAC)

The various particles will automatically organize themselves into clusters. These are small groups that are organized by certain features which are unique to all other clusters. This advance grading will separate unnecessary from desired information for the next steps.

Furthermore, a performance optimization of the DataOcean@ISAC visualization algorithm will be strived. The actual algorithm is limited to managing several thousand data objects, depending on the used hardware. The number of presentable particles has to be enhanced in order to use the visualization with big data structures. The use of JavaScript visualization runtime is browser based and adaptive to any display size. Therefore, the enhancement of the performance will guarantee proper usage of inefficient devices like tablets or smartphones in comparison to powerful PCs, touch tables or big touch walls. Figure 3 depicts an early version of the DataOcean@ISAC framework used on an 84" touch wall, which can be used in an authentic facility environment and is able to display big amounts of data.

Figure 3. Human Machine Interface: Using DataOcean@ISAC on an 84" touch wall

6. Conclusion and Outlook

The evaluation of current industry demands [agi05a] [agi05b] and national/international calls and guidelines [bmb15a] shows, that analysis, organization and evaluation of big pools of data objects are desired and needed, to improve the efficiency in several sectors of production and processing. ISAC@OTH-AW addresses these demands and will enhance the efficiency of production in SME. Besides the usage of well-known and proven algorithms, new types and advancements (e.g. interactive k-Means) will be developed.

Furthermore, current approaches of machine learning workflows are not based on interactive visualizations. ISAC@OTH-AW will implement new algorithms which will be based on visual analytics of the given data. This can be a promising approach.

The project started recently at the end of 2015 and will last till the end of 2021. This paper depicts the vision and intent of the project.

7. ACKNOWLEDGMENTS

8. REFERENCES

[Agi05a] Studie: Industrie 4.0 im Mittelstand: http://www.agiplan.de/unternehmen/news-und-presseinformationen/detail/studie-industrie-40-im-mittelstand.html

[Agi05b] Industrie 4.0 für den Mittelstand: http://industrie-40-mittelstand.agiplan.de/agiplan-industrie/?page=home

[Bin02a] Bin, W., and Yi Z. (eds.). Evolutionary Computation: CSIM: A document clustering algorithm based on swarm intelligence, in Conf.proc. CEC'02 Honolulu, HI, IEEE, pp.477-482, 2002.

[Bmb15a] Industrie 4.0 – Innovationen für die Produktion von morgen http://www.bmbf.de/pub/Industrie_4.0.pdf

[Cui06a] Cui, X., and Potok T.E. A Distributed Agent Implementation of Multiple Species Flocking Model for Document Partitioning Clustering. CIA 2006, LNAI 4149, pp.124-137, 2006.

[Cui05b] Cui, X., and Potok T.E. Swarm Intelligence Symposium: Document clustering using particle swarm optimization, in Conf.proc. SIS 2005, IEEE, pp.185-191, 2005.

[Cui09c] Cui, X., and Potok T.E. Swarm intelligence in text document clustering, in Handbook of research on text and web mining technologies, 2009.

[Fow12a] Fowler, R. H., Huerta, R. A. (eds.). Visualization and Clustering of Document Collections using a Flock-based Swarm Intelligence Technique, in Proceedings of the International Conference on Modeling, Simulation and Visualization Methods (MSV). The Steering Committee of the World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2012

[Eri11a] Erisoglu, M. and Calis, N. (eds.). A new algorithm for initial cluster centers in k-Means algorithm, in Pattern Recognition Letters 32, ScienceDirect, pp.1701-1705, 2011.

[Esm13a] Esmin, A.A.A. and Coelho, R.A (eds.). A review on particle swarm optimization algorithm and its variants to clustering high-dimensional data, in Artif. Intell. Rev, 44, Springer Science+Business Media Dordrecht, pp.23-45, 2013.

[Hart79a] Hartigan, J.A. and Wong, M.A. A k-Means Clustering Algorithm, in Journal of the Royal Statistical Society, Series C (Applied Statistics) 28, pp.100–108, 1979.

[Joll02a] Jolliffe, I. Principal component analysis. John Wiley & Sons, Ldt, 2002.

[Kuo11a] Kuo, R. J., and Wang, M. J. (eds.). An application of particle swarm optimization algorithm to clustering analysis, Soft Computing, 15(3), pp.533-542, 2011.

[Mal15a] Maltese, J., and Ombuki-Berman, B. (eds.). Co-operative Vector-Evaluated Particle Swarm Optimization for Multi-Objective Optimization, in Conf.proc. Computational Intelligence 2015, IEEE Symposium Series on IEEE, 2015.

[Mei15a] Meiller, D. Diving into the Data Ocean, International Federation for Information Processing 2015, Interact 2015, Part IV, LNCS 9299, pp.465-468, 2015.

[Na15a] Na, Wenbo, (eds.). Research on well production prediction based on improved extreme learning machine, in International Journal of Modelling, Identification and Control, pp.238-247, 2015.

[Sha12a] Shafeeq, A., and Hareesha, K.S. Dynamic Clustering of Data with Modified k-Means Algorithm, International Conference on Information and Computer Networks (CICIN 2012), IPCSIT vol.27, pp.221-225,2012